Does author team size explain retraction?

Peter Dieter
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

Retractions are a great concern in science, because researchers infringe the validity of their, but also the validity of other studies. Investigations showed that retraction rates rose significantly in the last years from about 0.01% in 2006 to 0.02% in 2012. However, only a few studies investigated this phenomenon so far. The purpose of this study therefore is to investigate retractions further. This will be done by looking at the number of researchers who contributed to a research article. Distinguishing between two reasons of retraction, namely errors of omission and errors of commission, applicable literature is reviewed. Non-retracted articles are then compared to retracted ones, in order to investigate a relationship between team size and errors of omission or errors of commission. The study shows that out of 450 retracted articles, 210 (47%) were retracted because of errors of commission, 97 (21%) because of errors of omission and for 143 (32%) articles, the reason of retraction could not been found, has been different, or the entries were corrections of articles already in the dataset. Research articles in the field of social sciences are more often retracted because of errors of commission (79%), than retracted research articles in Physics (66%) or Medicine (62%). Differently than hypothesized, the study shows that team size has little influence on the odds of an article on being retracted.

Graduation Committee members:

Dr. Tom De Schryver, Dr. A. H. (Rik) van Reekum

Keywords

Retraction, Team Sizes, Team Performance, Error of Commission, Error of Omission, Ethical Decision Making, Bibliometrics

9th IBA Bachelor Thesis Conference, July 5th, 2017, Enschede, The Netherlands.

1. INTRODUCTION

Before research papers are being published in research journals, they usually go through a diligent peer review session, in which the validity of the findings and methods are being checked. Publishers have the interest to conduct this process as accurate as possible, in order to achieve a good reputation, by assuring only high-quality papers being published in their research journals. Therefore, a publication is a reward and success for researchers. Despite this procedure, sometimes, errors are detected in papers after publication. In case of minor errors, such as an incorrect correspondence address or correction of an author's name, an erratum notice can be issued (Kulkarni, 2014). However, if the errors endanger the validity of the paper, it needs to be retracted. As Kulkarni (2014) explains: "Retraction is a way of alerting readers to the questionable credibility of a study". Often, authors themselves retract their articles, but also, journals, editors or institutions do retract (Wager & Williams, 2011).

In the research field of legal science, Haeberle (1957) states that there are two main errors individuals and groups can make when they are acting or deciding on something and introduced the terms "errors of omission" and "errors of commission". These terms can also be applied to other fields including scientometrics and retractions. Errors of omission are errors that have not been made consciously, such as a miscommunication or the usage of a wrong statistical test. Errors of commission however, are errors that were made with the researcher being conscious about the error. Typical examples here for are plagiarism or data fabrication. Both types of error are strongly imperiling the credibility and quality of studies.

Retraction rates rose significantly in the previous years. In 2006, the retraction rate in the "Web of Science" was below 0.01% and has more than doubled to around 0.02% in 2012 (Flanelli, 2013). Although past studies showed that this is mainly due to quicker detections of errors (Steen, Casadevall, & Fang, 2013), it is time to get a deeper understanding of retractions. Retractions are a big issue in the field of science, because researchers do not only infringe the validity of their own research, but also the validity of research articles that refer to these retracted articles. For example, according to retractionwatch.com, the most cited retracted paper (Voinnet, Rivas, Mestre, & Baulcombe, 2003) had 897 cites before it was retracted (retractionwatch.com, 2017). As it has been found that half of the retractions are due to fraud (Steen, Casadevall, & Fang, 2013), the rise of retraction rates does also provoke ethical issues in the field of science.

Studies already investigated retractions from different angles. A study by Lu et al. (2013) showed that retraction rates differ between different research disciplines, finding that retraction rates in "hard sciences" such as Biology or Chemistry, are higher than in "soft sciences" such as social science (Lu, Uzzi, Jones, & Jin, 2013). This may "reflect lower incidence of false science or lower rates of detection, where replication norms may differ" (Lu, Uzzi, Jones, & Jin, 2013). Another relationship that was discovered, is the association between retractions and countries, finding that countries such as China, India and South-Korea, have higher retraction rates than other countries (He, 2013). He (2013) concluded that this is due to inattentive research policies in these countries. Both studies did this by analyzing retractions rates in the Web of Science between 2000 and 2010.

In the last five decades, work-groups have become more dominant in the production of knowledge (Wuchty, Johns, & Uzzi, 2007). It was hypothesized, that this is mostly due to the increasing scale, complexity and cost in "hard sciences" (de Solla Price, 1963). However, this trend can also be observed in social sciences, where the just mentioned drivers (increasing scale, complexity and costs) "are much less notable" (Wuchty, Johns, & Uzzi, 2007). This trend might also influence scientific publications. Risks of errors could change depending on whether more or less researchers are contributing to a research article. Therefore, it is interesting to investigate which effects this development might bring in terms of retractions. In order to research the hypothetical relationship between number of authors and retraction rates, the following research question is proposed.

RQ: Do the odds of retraction change, if the number of authors per research article differs?

Previous research in the field of bibliometrics already researched the relationship between research team size and productivity, measured by quantitative variables: publications and citations (Cook, Grange, & Eyre-Walker, 2015) (Abt, 2017). These studies add on Lotka's law, which states that the number of authors who published n times, is decreasing exponentially with the number of publications (Lotka, 1926). Many authors have only one or a few contributions and less authors account for many contributions. This is expected to be mainly due to task division between researchers. A few supervising researchers who are on a high hierarchical level, appear in the authors list of many research articles. Most researchers however, are only contributing to few publications. Besides that, Cook et al. (2015), found a weak but significant positive relationship between group size and number of publications in the field of life sciences, by conducting a linear regression analysis with data from the Web of Science (WOS). Larger teams seem to be more productive because they have more resources (especially human resources). However, "the relationship is one of diminishing returns; productivity increases with funding but not proportionally" (Cook, Grange, & Eyre-Walker, 2015). These finding are complemented by the research of Abt (2017). By researching data in the field of astrophysics, he found evidence that research articles from larger teams are cited more often than articles from single authors or smaller teams, however, this is mostly due to higher self-citation of the involved authors.

Instead of analyzing the connection between research team size and performance from an angle that focuses on rather possible "positive" outcomes which deal more with productivity, namely by looking at publications and citations (Cook, Grange, & Eyre-Walker, 2015) (Abt, 2017), this study tries to analyze the relationship from another angle, namely the negative event of retractions. To this point of time, the relationship between number of authors and retractions has not been investigated. Researching this relationship, can enhance our understanding of possible factors that could influence the odds of an article being retracted, by providing new insights from a new point of view, namely by looking at author team sizes.

2. THEORETICAL FRAMEWORK

As already mentioned, in this study, it will be distinguished between errors of omission and errors of commission. These terms are used because they are classifying errors on a fundamental level, by distinguishing between unconsciously made (omission) and consciously made errors (commission). As these two errors have distinct reasons and are of different nature, for each reason, different literature will be reviewed in order to derive to proper hypotheses.

2.1 Errors of Omission

Articles that were retracted because of errors of omission, imply an inferior performance of the contributing researchers. Therefore, research in the field of organizational behavior and psychology, which investigated team size and their performance, will be reviewed.

Many studies have shown that team size has several consequences for the performance of an individual, but also for the overall team performance. Even though individuals in teams with more members have access to more nonphysical and physical resources, Hare (1952) also found out that individuals in larger groups participate less frequently because they feel that their opinion is not important.

Another reason for a worse performance is the finding, that individuals in larger teams assume less responsibility for the tasks which they are performing (Wicker & Mehler, 1971). More recent studies revealed that individuals in larger teams generally perform worse than individuals in smaller teams, however, overall group performance increases with team size (Liden, Wayne, Jaworski, & Bennett, 2004). An important reason for the worse performance of individuals in large teams, is the so called "social loafing", which can be explained as: exerting less effort when working in a group, as when working alone (Liden, Wayne, Jaworski, & Bennett, 2004). This research by Liden et al showed that social loafing is positively related to team size and negatively related to team performance.
Another explanation for the decreasing performance of individuals in larger teams was given by James Shepperd (1993). In his framework, he explains social loafing by introducing expectancy theory. An individual's motivation and performance decline when the individual does not see a link between their effort and reward. This link between effort and reward is often weaker in large work groups and thus, motivation and performance of individuals decline (Shepperd, 1993).

Other studies that investigated team size and overall group performance hypothesized a non-linear relationship, or more specifically an inversed U-shaped relationship, between these two variables. (Goodman, E., & Argote, 1986) (Barrasa, West, & Gil, 2007). Barrasa et al. (2007) conducted a study with health-care teams, in which evidence for this hypothesis was found. Team performance increased with team size, but only to a certain point, where performance started to decrease again.

A survey study by Krasnova et al. (2012) on research teams in the field of Information Systems, found that the first author contributes to around 60% of the work on a three-author paper, with the third author contributing only 15 percent of the work. Results of the same nature were expected in larger research teams. Even though the experience and knowledge of an additional author may improve the overall quality of the research paper, by providing new insights and feedback, it also comes with coordination and control costs and the lack of clarity regarding the attribution of credit (Krasnova, 2012). There might be a certain point until the contribution and knowledge of an additional researcher outweighs the negative effects of coordination and control costs and also lack of clarity. However, after this point is reached, it might be that the costs of adding an additional member are higher than the benefits.

This reasoning is closely related to the concept of "span of control". Many studies in the field of organizational behavior found out that team leaders are not able to coordinate and monitor the work of teams with too many members (Davis, 1951) (Van Fleet, 1977). Additional members are therefore not an advantage and come with higher costs than benefits.

2.1.1 Implications for team size and retractions

It is drawn on social psychological theories to make propositions about the relationship between X, the number of authors, and Y, retracted articles. Retractions have a high impact on the author's and co-authors' careers (Mongeon & Lariviere, 2016) and therefore, a high degree of responsibility is requested, namely the responsibility to conduct a valid research without any errors. As it is expected that that individuals take less responsibility in larger groups it can also be expected, that individual researchers feel less responsible to conduct a valid research without any error of omission in larger teams.

As it is still the tradition in research to emphasize the role of individuals in scientific discovery (Wuchty, Johns, & Uzzi, 2007), individuals in research teams might perceive that researching in a large team offers less rewards than conducting research alone or in smaller teams. This weak link between effort and reward is likely to result in less motivation and a weaker performance (Shepperd, 1993). This inferior performance in turn, could lead to higher odds for retractions. Besides that, with increasing team size, social loafing might increase with the effect that individual researchers would put less effort in their research, resulting in higher number of retractions.

However, other research found out that overall team performance is said to increase with team size (Liden, Wayne, Jaworski, & Bennett, 2004), and therefore the effect on retractions could also be ambiguity or non-linear. This reasoning is supported by the findings of Barrasa et al. (2007), who found an inversed U-shape relationship between team size and team performance. This leads to the hypothesis of a mirrored inversed U-shape relationship, or simpler spoken, a U-shape relationship between number of authors and retractions.

Besides that, based on the study of Krasnova et al. (2012), it can be assumed that there is point at which adding additional members results in higher costs of coordination and control than the benefits of an increase in knowledge and feedback would be. A loss of coordination and especially control mechanisms such as feedback loops, is likely to result in inferior performance and errors of teams. This is supported by the concept of "span of control" (Van Fleet, 1977) (Davis, 1951). There is a maximum number of team members the leading researcher can effectively coordinate and monitor.

These two findings are strengthening the hypothesis of a U-shape relationship between number of authors per article and retractions. Retractions because of errors of omission might first decrease with the number of authors until a certain point on which retractions because of errors of omission increase again. Therefore, the following hypothesis is proposed:

H1: There is a U-shaped relationship between the number of authors and retractions because of errors of omission.

2.2 Errors of commission

Articles that were retracted because of errors of commission, imply an unethical behavior of at least one of the involved researchers.

Many studies examined ethical decision making and its influencing factors on an individual level (Trevino & Youngblood, 1990) (Rest, 1979) (Choudhury, Mishra, Guyot, Meier, & Bell, 2012), finding that the individual decision making is influenced by two main features, namely demographic (e.g. age and income) and psychological (e.g. individuals' cognitive processes and locus of control) factors. However, less studies investigated the effects of groups on ethical decision making and cheating (O'Leary & Pangemanan, 2007) (Nichols & Day, 1982) (McCabe & Trevino, 1993) (Abdolmohammadi & Reeves, 2003).

Nichols and Day (1982), conducted an early study on 82 business students individually and in groups. They confirmed their hypothesis that "the level of moral judgment (…) is higher in the interacting group than the nominal average of the members comprising the group". Thus, groups have a higher level of moral judgement than their individual members. Evidence was found, that individuals who scored high (a high score implies an elevated level of ethical judgment), changed their behavior less in a team than individuals who scored low or average and thus, presumably influenced the group decisions more. The results of Nichols and Day (1982) raised the question, which other determinants influence the moral judgment, especially within a group. McCabe and Trevino (1993) found that even stronger than communicated ethical codes, the peer behavior influences individual's ethical decision making within a group. The effect is ambiguous though. Groups can create a social norm, in which misconduct is rather frowned upon but sometimes also supported. Next to that, a strong relationship between the certainty of being reported, and peer behavior was found (McCabe & Trevino, 1993). Therefore, it is concluded, that the certainty of being reported has a strong and negative influence on academic dishonesty.

There is also research that did not find any differences between the ethical decision making of individuals and groups. Abdolmohammadi and Reeves (2003), conducted a study with 98 students who followed the "Corporate Social Responsibility" at their institution. Students were first surveyed on an individual level and then sorted into three kind of groups: female only, man only and mixed groups. Differently than hypothesized, Abdolmohammadi and Reeves did not find significant differences between individuals and groups in any of these three kind of groups. However, the authors point out that their results should be interpreted cautiously, because "Corporate Social Responsibility" is an elective course at their institution and thus, there may be a self-selection bias in their data. Students taking this elective might already have high ethical standards when measured on an individual level.

O'Leary and Pangemanan (2007) investigated the behavior of 165 final year accountancy students, once in groups and once as individuals. It was found out that individuals are taking significantly more often the "extreme actions" of acting either unethical or ethical. Groups on the other hand, decide themselves more often for neutral decisions. "Groups appeared to reach a more consensus/compromised decision, most probably due to the increasing pressure to agree with others."

Another finding was provided by Zimbardo et al. (2003). By investigating the behavior of 576 psychology students who had to write an exam individually and in groups, he found enough evidence that students cheat less often in larger groups, than when writing the exam alone.

2.2.1 Implications for team size and retractions

Even though not all studies concluded that larger groups act more ethical than individuals (Abdolmohammadi & Reeves, 2003), there is a basis for expecting that individuals conduct errors of commission more often than groups. As Nichols & Day (1982) showed that groups have a higher standard of moral judgment than their individual members and errors of commission are a heavy breach of moral codes, it can be expected that individuals conduct errors of commission more often than groups. Also, Zimbardo et al. (2003), found evidence that larger groups cheat less often. Besides that, McCabe and Trevino (1993) found out that the certainty of being reported has a negative relationship on academic dishonesty, and as it can be expected that the certainty of being reported increases with a higher number of group members, also a negative relationship between the number of group members and errors of commission is expected. Another angle, which focuses more on the individual level is provided by O'Leary and Pangemanan (2007). Individuals take more often the extreme actions, including an unethical behavior and hence can also be expected to conduct errors of commission more often. By incorporating these previous researches in can be expected that larger groups are expected to conduct errors of commission less often than smaller groups or individuals and thus the following hypothesis is proposed:

H2: There is a negative (linear) relationship between articles that were retracted because of errors of commission and group size.

3. METHODOLOGY

As this research is built upon a hypothetico deductive reasoning, in this section, procedures will be presented that are used in order to confirm or reject the research hypotheses and by doing that, providing an answer to the research question. This will be done by explaining the relevant variables, as well as the applied sampling strategy and statistical tests.
Research articles listed in the Web of Science (WoS) (by Thomson Reuters), will be considered for conducting statistical tests. The advantage of the Web of Science lies in the high standards with which articles are evaluated. The Thomson Scientific Editorial Development group "carefully evaluates journals for potential inclusion in the database" (Thomson, 2017). The objective of the WOS is "to include only the most influential, relevant, and credible journal information available" (Thomson, 2017). Besides that, retracted articles in the data base are marked as such.

3.1 Sampling

Past studies have already shown an association between retraction rates and research disciplines (Lu, Uzzi, Jones, & Jin, 2013), finding that retraction rates are higher in "hard sciences" than in "soft sciences". Besides that, group sizes differ among research disciplines, being significantly smaller in social sciences than in other fields of research (Wuchty, Johns, & Uzzi, 2007). Thus, it is important to take the research discipline into account when sampling. Another research by He (2013) has shown that retraction rates are higher in countries such as China, India or South-Korea, than in western countries. Even though, it is questionable if group sizes differ significantly among countries, they are another factor that can be considered when sampling. Taking the above-mentioned findings into account, the sampling strategy that is applied is a stratified sampling which is based upon countries and research disciplines. To reduce complexity of the study and obtain a representative sample, the research only focuses on three research disciplines of different nature, namely Biology and Medicine, Physics and Social Sciences. Besides that, only three continents will be taken into account: Asia, America and Europe. Because of a limited population of retracted articles within fields of study (e.g. social science) in specific countries (e.g. China), the research in Asia contains articles from China and South-Korea. The same applies in Europe, where research articles from the UK and Germany will be considered. In America however, only articles from the United states will be taken into account.
Per specific continent and discipline, 50 retracted and non-retracted articles are sampled. Thus, per continent, 150 retracted and 150 non-retracted research articles are sampled. A graphical sampling scheme is to be found in appendix 1. In order to assure randomness, not only the first 50 articles displayed in the Web of Science within each subset will be downloaded, but a random generator determines ten blocks of five research articles which will be taken into account.

3.2 Variables

The dependent variable "retraction" is a categorical variable that can only have the values "Non-Retracted", "Commission" or "Omission". To find all retracted articles in the WoS, the following search terms are used: the title of the articles contains either "retraction AND VOL" or "retracted AND article". This search query has already been applied by He (2013). Within retracted articles it is distinguished between articles that have been retracted because of errors of commission and research articles that have been retracted because of qualitative or communicational errors within the research team (errors of omission). This information is available in the retraction notice and has to be checked manually. The independent variable on which is focused on, is the number of authors (NOA). It is a numerical variable that can theoretically range from one to infinity. The WoS does not provide this information, but only states the author names, which are separated by a semicolon. Therefore, an Excel formula will be applied in order to extract the variable "number of authors":=(Len(Reference)-Len (Substitute(Reference,";","")) +1. Other independent variables that are considered are the continent and research discipline. These two variables are both categorical and can only have the values Asia, Europe, America and Medicine, Physics and Social, respectively. Thus, they will be dummy coded. Next to that, the interval variable "publication year" and the ratio variable "page numbers" will be included, in order to increase the completeness of the model and to account for possible intercorrelation effects.

3.3 Statistical tests

For each hypothesis, a statistical test will be applied. This test is in both cases a binomial logistical regression, as the dependent variable can only have two values per case, namely non-retracted or, depending on the case, either retracted because of errors of commission or retracted because of errors of omission. Retracted articles are defined as 1, non-retracted articles defined as 0. For every test, all 450 non-retracted articles serve as a sample for articles that have not been retracted. For hypothesis one, these 450 articles are then compared only to articles retracted because of errors of omission. For hypothesis two, these 450 articles are compared only to articles that are retracted because of errors of commission.

The first hypothesis, which assumes a non-linear relationship between number of authors and retraction because of errors of omission, will be tested using the following logistical regression model:

$$Logit(p_{Omission}) = \beta_0 + \beta_1\, number\ of\ authors + \beta_2\, number\ of\ authors^2 + \beta_3\, Year + \beta_4\ \ number\ of\ pages + \beta_5 Research\ Discipline + \beta_6\, Continent + \varepsilon$$

Here, we expect the coefficient β_2 to be positive, as we hypothesize a U-shaped relationship.

The second hypothesis, which assumes a negative and linear relationship between the number of authors and retraction because of errors of commission, will be tested using the following model. The coefficient β_1 is expected to be negative.

$$Logit(pc_{ommission}) = \beta_0 + \beta_1\, number\ of\ authors + \beta_3\, Year + \beta_4\ \ number\ of\ pages + \beta_5 Research\ Discipline + \beta_6\, Continent + \varepsilon$$

4. RESULTS

4.1 Descriptive statistics

The sample revealed that 97 articles have been retracted because of errors of omission, 210 because of errors of commission and for 143 articles, the reason of retraction could not been found, has been different, or the entries were corrections of articles already in the dataset. The "commission-omission ratio" was highest in social sciences, with around 79%, followed by Physics (66%) and Medicine (62%). This means that retracted research articles in the field of social sciences are more often retracted because of errors of commission, than retracted research articles in Physics or Medicine. Besides that, the "commission-omission ratio" was highest in Asia (75%), followed by Europe (66%) and America (63%). These findings are summarized in a contingency table below.

Table 1: Descriptive statistics

	Commission	Omission	N.A.	Tot.
Medicine	74 (16%)	45 (10%)	31 (7%)	150
Physics	67 (15%)	34 (8%)	49 (11%)	150
Social	69 (15%)	18 (4%)	63 (14%)	150
Total	210	97	143	450
Europe	69 (15%)	35 (8%)	46 (10%)	150
America	65 (14%)	37 (8%)	48 (11%)	150
Asia	76 (17%)	25 (6%)	49 (11%)	150
Total	210	97	143	450

Next to that, the previous research findings of Wuchty et al. (2007) could be confirmed. Research team sizes differ significantly among research disciplines and are indeed smaller in social sciences and largest in Medicine. Excluding two influential outliers in the field of Physics with 428 and 367 authors, the following violinplot (Figure 1) can be drawn. The black point represents the median.

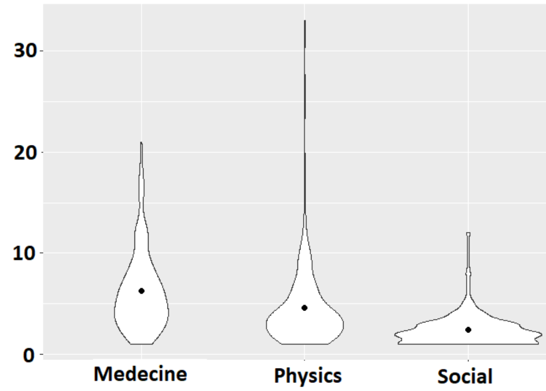

Figure 1: Violin: Number of Authors (NOA) ~ Research Discipline

Including these two extreme points, the mean number of authors in Physics is highest with 7.23 authors. Excluding, the two outliers, the author mean in Medicine is highest with 6.23 authors, followed by Physics with 4.61 authors, and Social Science with 2.4 authors per article. Even when the two outliers in the field of Physics are excluded, the standard deviation is still highest in Physics with 4.1 authors, followed by Medicine (3.57) and Social Science (1.7). A one-sided ANOVA confirms that these findings are significant with a p-value of 0.00426 (output in Appendix 2). Conducting the same test with continents and countries instead of research disciplines reveals that group sizes do not differ significantly among continents, with a p-value of 0.06 (output in Appendix 2). This finding can be visualized with the following violinplot (Figure 2).

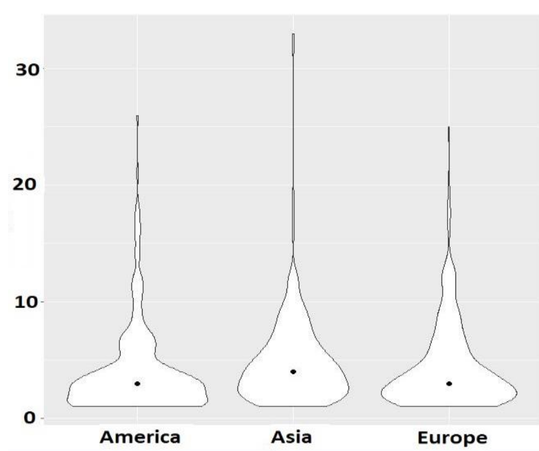

Figure 2: Violin: Number of Authors (NOA) ~ Continent

Further statistics (excluding the two outliers in the field of physics) including the mean, median, standard deviation, minimum and maximum, are summarized in the table below.

Table 2: Comparison of author means, medians and standard deviation by Research Discipline and Continent

	Mean	Median	Std. Dev	Min	Max
Medicine	6.23	5	3.91	1	21
Physics	4.61	4	4.1	1	33
Social	2.4	2	1.7	1	12
Europe	4.38	3	3.67	1	25
America	4.06	3	3.8	1	26
Asia	4.79	4	3.76	1	33

4.1.1 Correlation Matrix

Next to these descriptive statistics, a deeper analysis on some variables is conducted. Below, a correlation matrix (Table 3) with the most important numerical variables is displayed. The data used for this matrix were all 900 articles. It was not separated between retracted and non-retracted articles. It shows some interesting relationships.

Table 3: Correlation matrix

	Times Cited	Year	# of Pages	# of Authors
Times Cited	1			
Year	-0.096**	1		
# of Pages	0.147	-0.033	1	
# of Authors	0.055	0.1**	-0.139***	1

*** Significant at 0.001 level (two-tailed)
**Significant at 0.01 level (two-tailed)

First of all, the findings of de Solla Price (1963) and Wuchty et al. (2007) that team sizes increased in the last decades can be confirmed. The coefficient of 0.1 is positive, meaning that the model states that the average team size increased by 0.1 authors per year, in the last decades. This relationship is significant at a 0.01 level (two-tailed). Besides that, the findings of Abt (2017), that articles that were written by more authors are also cited more often than articles written by less authors, could not be confirmed. Even though there is a positive relationship between team size and the times a paper was cited, the correlation is not significant Another interesting finding is that more recent articles are less often cited (Coefficient = -0.0962, two tailed p-value = 0.0041). This might have several reasons. Older research articles can be more often cited just because of the fact that they are longer published than more recent ones. Besides that, the number of articles published increased a lot over time, and the fewer articles that are published earlier, are therefore more likely to act as the foundation of more recent articles.
The last significant relationship in the correlation matrix seems quite surprising. The more authors contributed to an article, the less pages the article has. However, there is a logical explanation for that. Research articles in social sciences have less authors, but a higher number of pages. Research articles in medicine have more authors and less pages. Therefore, when all research articles are aggregated, a negative correlation between those two variables appears. Another analysis in which research disciplines are separated, reveals a positive, but insignificant relationship between the number of authors and the number of pages for all three research disciplines.

4.2 Hypothesis

In this chapter, the hypotheses which were presented in chapter two, are tested. The findings are summarized in a separate table for each hypothesis. These tables include coefficients of the variables, state if they are significant correlated and contain as well a Hosmer-Lemeshow and McFadden (Pseudo $R^{2)}$ test, which assess the goodness of fit of the models or in other words, whether the observed data values match the expected ones. Both values can range from 0 to 1. The lower these values are, the worse the model is predicting.

4.2.1 Errors of Omission

The first hypothesis, which assumes a positive U-shaped relationship between retraction and number of authors, reveals interesting results. Surprisingly, the coefficient β_2 is significantly negative. The test results are summarized in the table below.

Table 4: Summary: Regression for hypothesis 1

Model	Baseline Model	Linear Model	Full Model
Intercept	-183 ***	-181***	-173***
Physics	-0.35	-0.33	-0.23
Social	-0.59	-0.54	-0.19
Asia	-1.16***	-1.16***	-1.30***
Europe	-0.25	-0.26	-0.33
Year	0.09***	0.09***	0.09***
# of Pages	-0.12***	-0.12***	-0.12***
NOA		-0.03	0.27*
NOA^2			-0.01*
Hosmer-Lemeshow	0.0089	0.045	0.087
McFadden test	0.1309	0.131	0.145

p-value in brackets
*** Significant at 0.001 level (two-tailed)
**Significant at 0.01 level (two-tailed)

Including the number of author variable, the following formula is obtained:

Logit($p_{Omission}$)= -173 + 0.27 number of authors – 0.01 number of authors^2 – 0.23 Physics – 0.19 Social – 1.3 Asia - 0.33 Europe -0.12 number of pages+ 0.09 Year + ε

The p=value of β_2 (number of authors^2) is significant with a value of 0.022. This finding is quite surprising because the opposite direction was hypothesized. This would mean that retractions because of errors of omission first increase with the number of authors until a certain point on which retractions because of errors of omission decrease again. Splitting the data up into the different research disciplines, show that this negative relationship can be found in all research disciplines, however it is only significant in Social Science and by no means significant in Medicine (p=0.92) and Physics (p=0.24). Taking a closer look at the logistic regression plot of only articles in the field of social science, it can be observed that retractions are indeed first increasing, however, there is one non-retracted article with 11 authors, which is strongly influencing the regression outcome. When this data point is removed, another regression only in the field of social science reveals that the coefficient β_2 is becoming positive (0.031), but not significant (p=0.86).

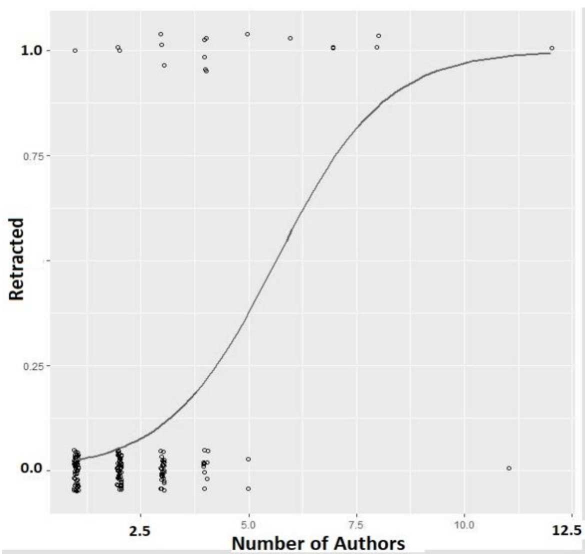

Figure 3: Logistic Regression Plot in Social Science Retracted ~ Number of Authors

Next to that, it can be observed that the dummy variable "Asia" is significantly negative related to errors of omission in all three models. This supports the descriptive statistics (Table 1), where it could be observed that errors of omission appear the least in Asia. Besides that, the logistic regression analyses revealed more relationships. The finding that retraction rates increased (Flanelli, 2013) can be confirmed. The publishing year correlates significantly with retractions, with a positive coefficient of 0.09 and a p-values lower than 0.001 in the full model. This implies that articles that are published more recently are more likely to get retracted than less recent ones. Next to that, a relationship between the number of pages and retractions was found. The coefficient here for is -0.12 and highly significant with a p-values lower than 0.001 in all three models. A further analysis shows that this observation can be made in all three research disciplines. Retracted articles seem to have more pages than non-retracted ones.

The goodness of fit measures both increased when the number of authors was added, namely from 0.089 to 0.045 in the linear model to 0.087 in the full model (Hosmer-Lemeshow) and from 0.1309 in the baseline model to 0.145 in the full model (McFadden), however, their values are still quite low. This implies that the overall model, including the number of authors, is not well predicting retractions because of errors of omission. This can also be observed when looking at the ROC curve, which is to be found in the appendix 3. Nevertheless, we cannot confirm the first hypothesis. There seems to be no positively U-shaped relationship between the number of authors and retractions because of errors of omission.

4.2.2 Errors of Commission

In contrast to the previous hypothesis, the second hypothesis focuses not on retracted articles because of errors of omission but because of errors of commission. The test results are summarized in table 5.

Table 5: Summary: Regression for hypothesis 2

Model	Baseline Model	Full Model
Intercept	-135***	-142.6***
Physics	-0.06	-0.13
Social	0.17	-0.03
Asia	-0.27	-0.26
Europe	0.06	0.08
Year	0.07***	0.07***
# of Pages	-0.05***	-0.05***
NOA		-0.05
Hosmer-Lemeshow	P<0.001	P<0.001
McFadden test	0.059	0.063

p-value in brackets
*** Significant at 0.001 level (two-tailed)
**Significant at 0.01 level (two-tailed)

As it is showed in table 5, the following logistical regression formula is obtained:

Logit($p_{Comission}$)= -142.6 – 0.05 number of authors – 0.13Physics – 0.03 Social -0.26 Asia + 0.08 Europe + 0.07 Year -0.05 number of pages ε

It can be observed that larger groups conduct errors of commission indeed less often than smaller groups or individuals. The sign of β_1 is negative as hypothesized, however, the relationship is not significant with a p-value of

0.062. Separating the analysis on the three-different research discipline reveals that significance is highest in Medicine (0.047), followed by Social Sciences (0.2) and Physics (0.26). Also here, the same findings regarding the year of publication and the number of pages were made. The year of publication is significantly positive related to retractions because of errors of commission with a coefficient of 0.07 and a p-value lower than 0.001 in the baseline model as well as in the full model. The number of pages is significantly negative related to retractions because of errors of commission with a coefficient of -0.05 and a p-value lower than 0.001.

The goodness of fit measures increased only to a small amount (measured by percentage change), when the number of authors was added as an independent variable. The Hosmer-Lemeshow value stayed below 0.001, the McFadden increased from 0.059 to 0.063. Also here, the goodness of fit values are very low. Looking at the ROC curve (Appendix 3), confirms these findings. Therefore, this model seems not predicting retractions well. Hence, we cannot confirm our second hypothesis due to lack of statistical significance.

5. DISCUSSION

5.1 Limitations

There are several limitations to be mentioned. Keeping them in mind, assures that we stay realistic and enhances our understanding about implications of the research. For example, it could be observed that especially in the field of social sciences and physics, single authors who committed errors of commission or conducted multiple studies with errors that were made unconsciously, account for a large amount of retractions within the sub-samples. An example here for is the German economist Ulrich Lichtenthaler. In the subsample of retracted articles in Europe within the field of social science, he accounted for 24% of all articles. Therefore, the sample is likely to lack independence. However, independence is a major assumption for logistical regression. This fact is certainly affecting the validity of this study negatively. Another key limitation is the restricted population size of sub-samples. For instance, only 18 research articles in the field of Social Science have been retracted because of errors of omission. This finding is important when statements are made about generalizations of the study, as it endangers the external validity.

Next to these methodological issues, there are also theoretical issues to mention. Mainly research in the field of psychology and sociology was applied to derive to proper hypotheses. This research rarely analyzed the dynamics in research teams, but focused more often on working teams in the industry or student groups. However, research teams might have different dynamics including their own rules and procedures, which distinguishes them from other teams. Therefore, the applied literature might, at least partially, not be a suitable fundament, to base valid hypotheses on. Besides that, it was assumed that all errors made in research articles have the same chance to get detected. However, there might also be factors influencing the detection of errors, which require a separate theoretical framework including distinct arguments and hypotheses.

5.2 Future Research

First, it is to mention that due to very low goodness of fit values, the applied models, including the number of authors, did only explain a small part of retractions. Even though increasing the sample size, could reduce this problem, this raises the question, which other factors do explain retractions. In order to answer this question, future research should conduct investigations on different research levels, as there are many possible factors, that could influence the odds of retraction. The logistical regression models in section 4.2 confirmed the finding of Flanelli (2013), that retractions rate rose in the last years. Also, the logistic regression models showed that the number of pages seems to explain a part of retractions. In both cases, errors of omission and commission, retracted articles seem to have more pages than non-retracted ones. So far, retractions were not researched from this angle and therefore, future research might investigate this finding further and give an explanation for it.

Additionally, factors that influence the detection probability of errors should be investigated. Also here, the number of authors could play a role in the detection of errors. Next to that, future research could address the peer review process and compare and analyze the processes within research institutes or research journals. An interesting approach could be the question if a peer-review process with many “peers” results in a better error detection. Another approach could be the comparison of different research institutes. Do for example, articles from researchers of top ranked universities, have lower odds of being retracted.

6. CONCLUSION

Even though increasing retraction rates can be interpreted as a good sign (Flanelli, 2013), there is no doubt about the fact, that retracted articles are a negative event which can have severe consequences. Understanding the reasons for retractions better, might help to diminish retractions and their severe consequences. Increasing retraction rates are a phenomenon in the field of bibliometrics, that have been investigated by previous studies (Flanelli, 2013) (He, 2013) (Lu, Uzzi, Jones, & Jin, 2013). These studies explained retractions on a country or research discipline level (He, 2013) (Lu, Uzzi, Jones, & Jin, 2013). The purpose of this study however, was to investigate retractions on a lower (research group) level, namely by looking if team sizes have an impact on the odds of retractions. This was done by splitting up 450 retracted articles into two groups, namely articles that have been retracted because of errors of omission and articles that have been retracted because of errors of commission, and comparing those in a binomial logistical regression to non-retracted articles. Differently than hypothesized, the analyses showed, that the number of authors have a rather small impact on the odds of an article on being retracted. Excluding influential data points, the relevant coefficients were not significant and goodness of fit measures were close to zero. The applied literature from psychological and sociological fields, might not be a suitable fundament to describe detected errors made in research teams that leaded to retractions. This raises the question, which other reasons might influence the odds of an article on being retracted. Other literature and factors which focus more on the detection of errors, such as the peer review process, the citation index of a researcher or the ranking of research institutes, should be considered for further investigations in order to get a deeper understanding of retractions.

7. ACKNOWLEDGEMENTS

Hereby I would like to thank everybody who helped me writing the thesis. Especially to mention is my first supervisor Tom de Schryver, with whom I held individual meetings regularly. I know and appreciate that this goes beyond his duties. Next to mention are my second supervisor Mr. Rik van Reekum, as well as Mr. Paul van de Vet and Mrs. Anja Rebber, who gave me valuable feedback on the theoretical framework as well as on methodological aspects.

BIBLIOGRAPHY

Abdolmohammadi, M. J., & Reeves, M. F. (2003). Does Group Reasoning Improve Ethical Reasoning? *Business and Society Review: Vol. 108 Issue 1*, 127-137.

Abt, H. (2017). Citations and Team Sizes. *Astronomical Society of the Pacific*.

Barrasa, A., West, A., & Gil, F. (2007). Is there an Optimal Size for Health-Care Teams? Effects on Team Climate for Innovation and Performance. In *Psychosocial Resources in Health Care Systems* (pp. 51-64). Hampp,Rainer.

Choudhury, D., Mishra, S., Guyot, W. M., Meier, R. J., & Bell, R. L. (2012). The impact of social and demographic variables on ethical decision making: Exploratory study. *International Journal of Business and Public Administration, Volume 9, Number 3*, 126-144.

Cook, I., Grange, S., & Eyre-Walker, A. (2015). Research Groups: How big should they be? *PeerJ*.

Davis, R. (1951). *The fundamentals of top management.* New York: Harper.

de Solla Price, D. (1963). Little Science, Big Science. *Columbia University Press*.

Flanelli, D. (2013). Why Growing Retractions Are (Mostly) a Good Sign.

Goodman, P., E., R., & Argote, L. (1986). Current Thinking about Groups: Setting the Stage for new IDeas. In *Designing Effective Work Group* (pp. 1-33). San Francisco: Jossey-Bass.

Haeberle, M. (1957). ERRORS OF OMISSION, COMMISSION. *AMERICAN BAR ASSOCIATION JOURNAL*, 876&.

He, T. (2013). Retraction of global scientific publications from 2001 to 2010. *Scientometrics*, 555-561.

Jennings, D. F., Hunt, T. G., & Munn, J. R. (1996). Ethical decision making: An extension to the group level. *Journal of Managerial Issues, 8(4)*, 425-439.

Krasnova, H. T.-F. (2012). Let's Collaborate, But I Will Be the First Author! Exploring the Importance of the First Authorship for IS Researchers. *European Conference on Information Systems.* Barcelona.

Kulkarni, S. (2014, 29 08). *editage.com.* Retrieved from http://www.editage.com/insights/what-every-researcher-should-know-about-retraction

Liden, R., Wayne, S., Jaworski, R., & Bennett, N. (2004). Social Loafing: A field investigation. *Journal of Management*, 285-304.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences. 16 (12)*, 317–324.

Lu, S., Uzzi, B., Jones, B., & Jin, G. (2013). The Retraction Penalty: Evidence from the Web of Science.

McCabe, D. L., & Trevino, L. K. (1993). Academic Dishonesty: Honor Codes and Other Contextual Influences. *The Journal of Higher Education, Vol. 64, No. 5*, 522-538.

Meltzoff, J. (1998). *Critical Thinking about Research: Psychology and Related Fields.* Amer Psychological Assn.

Mongeon, P., & Lariviere, V. (2016). Costly Collaboration: The Impact of Scientific Frau on Co-authors' Careers. *Journal of the Association for Information Science and Technology, 67*, 535-542 .

Nichols, M. L., & Day, V. E. (1982). A Comparison of Moral Reasoning of Groups and Individuals on the ‘‘Defining Issues Test’’. *Academy of Management Journal 25*, 201–208.

O'Leary, C., & Pangemanan, G. (2007). The Effect of Groupwork on Ethical Decision-Making of Accountancy Student. *Journal of Business Ethics*, 215-228.

Rest, J. R. (1979). Development in Judging Moral Issues. *University of Minnesota Press, Minneapolis*.

retractionwatch.com. (2017, March 17th). Retrieved from http://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/

Shepperd, J. (1993). Productivity Loss in Performance Groups: A Motivation Analysis. *Psychological Bulletin Vol. 113 No. 1*, 67-91.

Steen, G., Casadevall, A., & Fang, F. (2013). Why Has the Number of Scientific Retractions Increased?

Thomson. (2017, April 5). *Thomson.com.* Retrieved from http://wokinfo.com/media/pdf/wosbenefits_general.pdf

Trevino, L. K., & Youngblood, S. A. (1990). Bad Apples in Bad Barrels: A Causal Analysis. *Journal of Applied psychology Vol. 75*, 378-385.

Van Fleet, D. a. (1977). A history of the span of management. *Academy of Management Review*.

Voinnet O, R. S. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. . *Plant Journal*.

Wager, E., & Williams, P. (2011). . Why and how do journals retract articles? An analysis of Medline retractions 1988–2008. *Journal of Medical Ethics*, 567–570.

Wicker, A., & Mehler, A. (1971). Assimilation of new members in a large and a small chruch. *Journal of Applied Psychology* , 151-156.

Wuchty, S., Johns, B., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 1036-1039.

Zimbardo, P., Butler, L. D., & Wolfe, V. A. (2003). Cooperative College Examinations: More Gain, Less Pain When Students Share Information and Grades. *The Journal of Experimental Education*, 101-125.

APPENDIX

1. Methodology

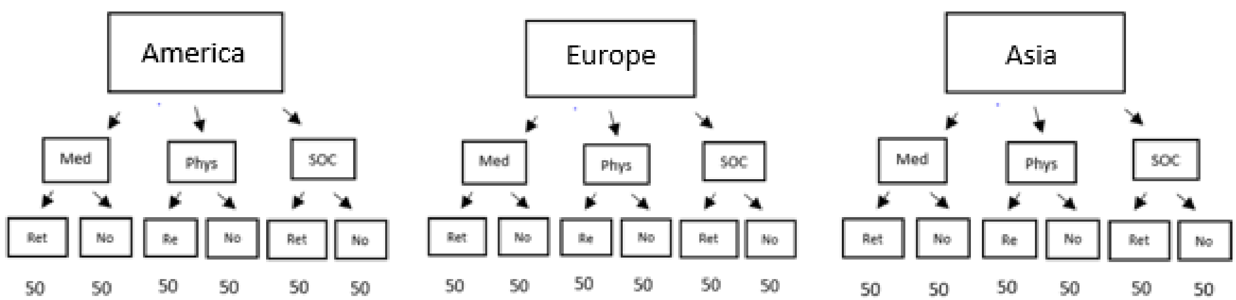

Figure 3: Sampling plot

2. Descriptive statistics

```
> fit = lm(formula = NOA ~ RD, data=data)
> anova(fit)
Analysis of Variance Table

Response: NOA
           Df  Sum Sq Mean Sq F value    Pr(>F)
RD          2  2221.8 1110.89  95.213 < 2.2e-16 ***
Residuals 895 10442.3   11.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: one-sided ANOVA (Number of Authors ~ Research Discipline)

```
> fit = lm(formula = NOA ~ Continent, data=data)
> anova(fit)
Analysis of Variance Table

Response: NOA
           Df  Sum Sq Mean Sq F value  Pr(>F)
Continent   2    79.9  39.967  2.8425 0.05881 .
Residuals 895 12584.1  14.060
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: one-sided ANOVA (Number of Authors ~ Continents)

3. Hypothesis

```
Call:
glm(formula = Retracted ~ RD + Continent + PY + PG + NOA + I(NOA^2),
    family = binomial, data = quality)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5143  -0.6559  -0.4473  -0.1810   2.6217

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.728e+02  3.995e+01  -4.326 1.52e-05 ***
RDPhysics       -2.262e-01  2.818e-01  -0.803   0.4220
RDSocial        -1.945e-01  3.776e-01  -0.515   0.6065
ContinentAsia   -1.300e+00  3.310e-01  -3.927 8.62e-05 ***
ContinentEurope -3.333e-01  2.894e-01  -1.152   0.2494
PY               8.573e-02  1.993e-02   4.301 1.70e-05 ***
PG              -1.244e-01  2.651e-02  -4.694 2.68e-06 ***
NOA              2.723e-01  1.101e-01   2.474   0.0134 *
I(NOA^2)        -1.479e-02  6.455e-03  -2.291   0.0220 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

Figure 5: Hypothesis 1. Full model. Key statistics output in R

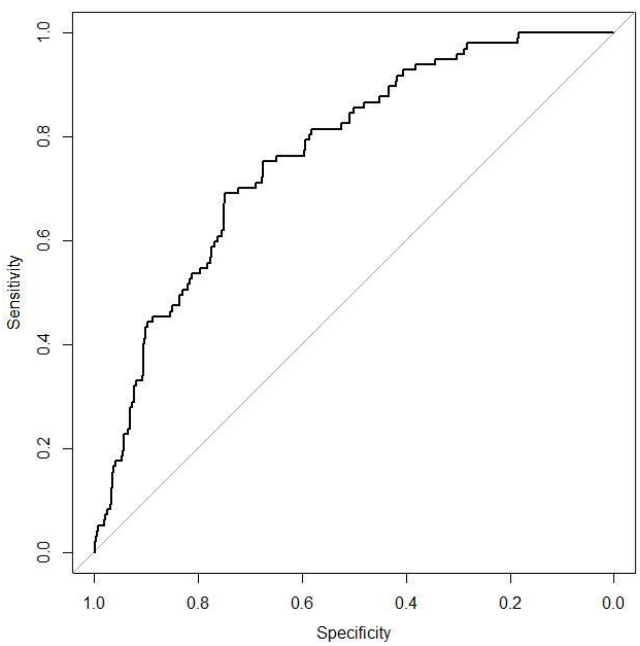

Figure 5: ROC curve: Retracted ~ NOA + NOA^2+ Research Discipline + Continent + number of Pages + Year

```
Call:
glm(formula = Retracted ~ RD + Continent + PG + PY + NOA, family = binomial,
    data = fraud)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3888  -0.9142  -0.6944   1.2052   2.2201

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -142.60752   27.76601  -5.136 2.81e-07 ***
RDPhysics         -0.13239    0.21550  -0.614    0.539
RDSocial          -0.03543    0.24800  -0.143    0.886
ContinentAsia     -0.26703    0.22351  -1.195    0.232
ContinentEurope    0.07531    0.21856   0.345    0.730
PG                -0.05245    0.01320  -3.972 7.11e-05 ***
PY                 0.07105    0.01385   5.130 2.90e-07 ***
NOA               -0.05178    0.02774  -1.866    0.062 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

Figure 6: Hypothesis 2. Full model. Key statistics output in R

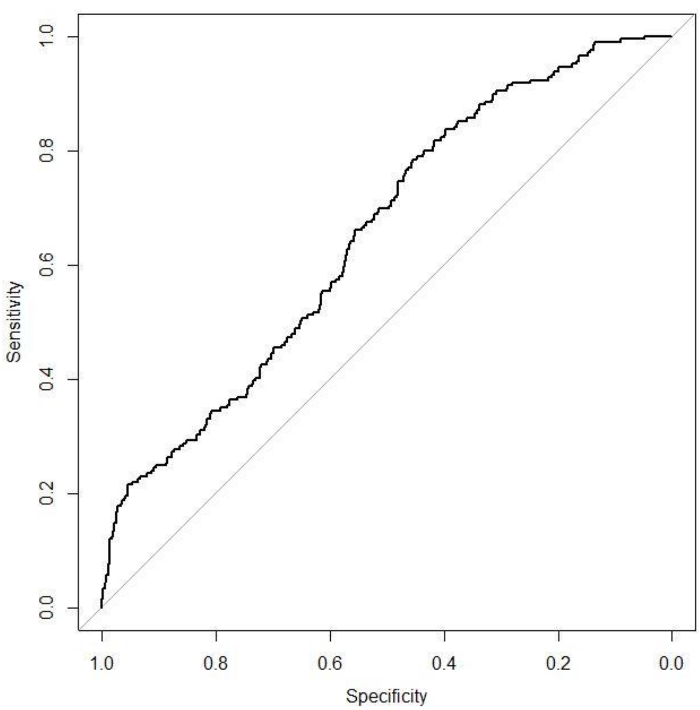

Figure 7: Hypothesis 2: ROC curve: Retracted ~ NOA + Research Discipline + Continent + number of Pages +Year